Introduction to Bioinformatics
6. **DNA Restriction Analysis and Primer Design**

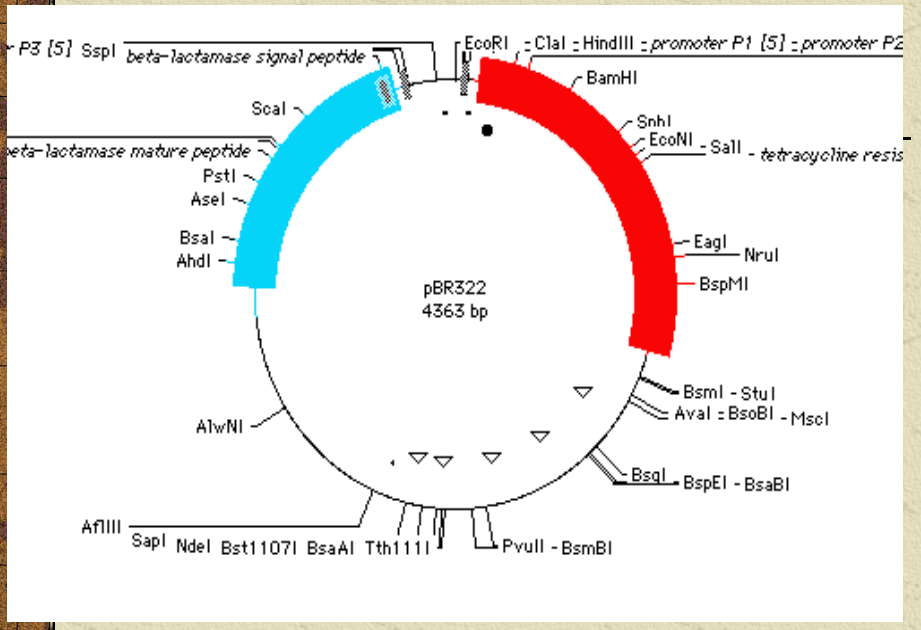
Benjamin F. Matthews
United States Department of Agriculture
Soybean Genomics and Improvement Laboratory
Beltsville, MD 20708
matthewb@ba.ars.usda.gov

What we will cover today

- Restriction site analysis
  - Cloning
  - Southern blot analysis
- PCR primer design
  - Cloning, amplification gene expression

Restriction mapping

- For cloning
- Southern blot interpretation



Restriction Mapping

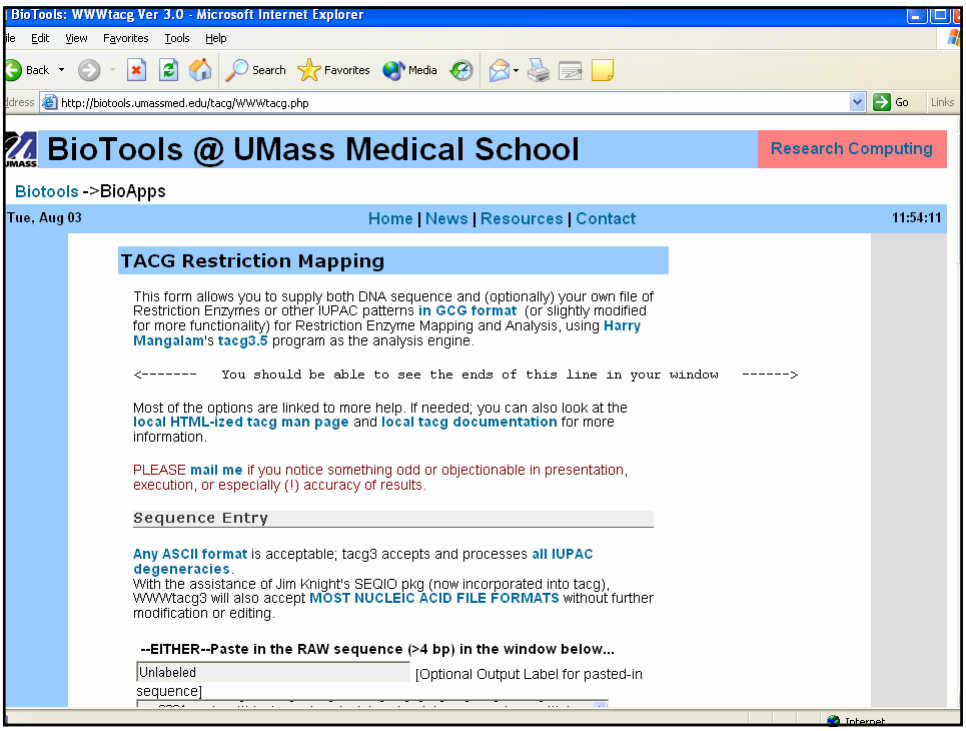
- Making restriction maps was my first use of "Molecular Biology" software
- Making restriction maps is a routine lab activity that is necessary for any type of cloning project.
- High quality maps are important for publications and exchange of information between researchers or between labs.

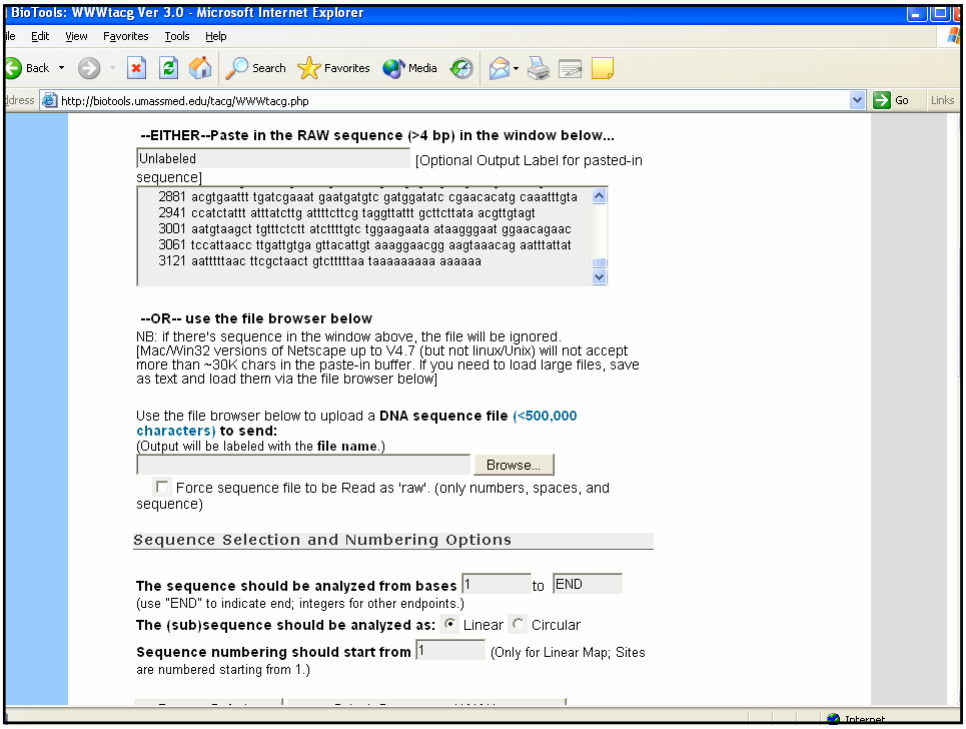
Mapping Software

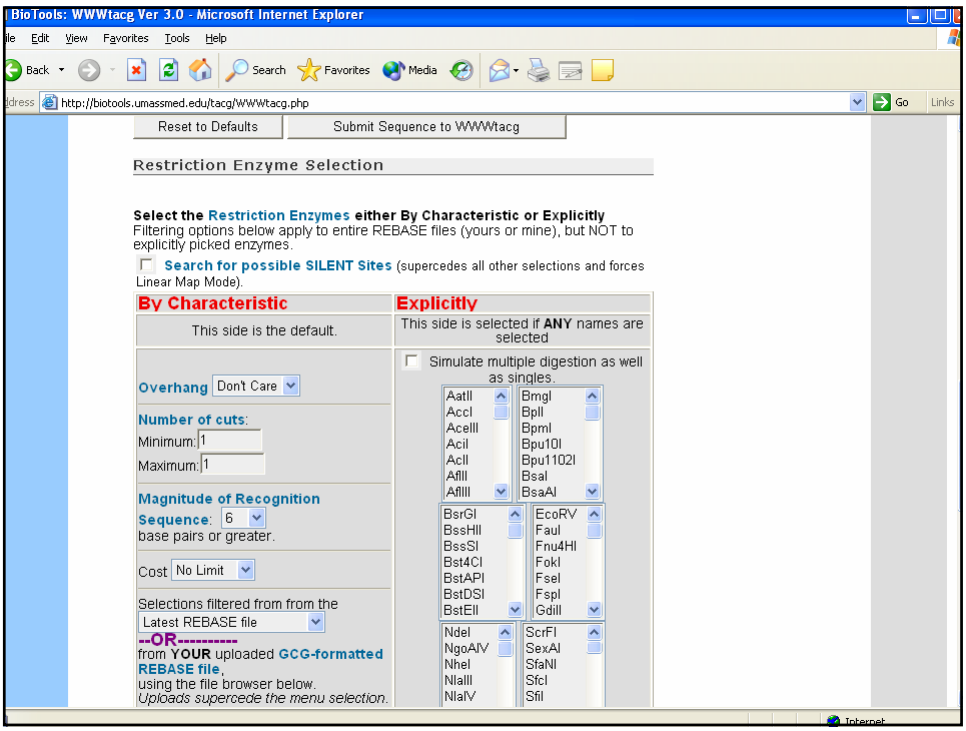
- Programs vary greatly in sophistication and ease of use
  - Simple drawing programs (vector graphics)
  - The venerable DNA Strider
  - GCG (not a strong point of the package)
  - Comprehensive Mac/PC MolBio programs
  - Dedicated plasmid drawing programs
  - Can it be done on the Web?
- Making high quality graphical restriction maps is one area where Mac/PC programs are much better than GCG or the Web

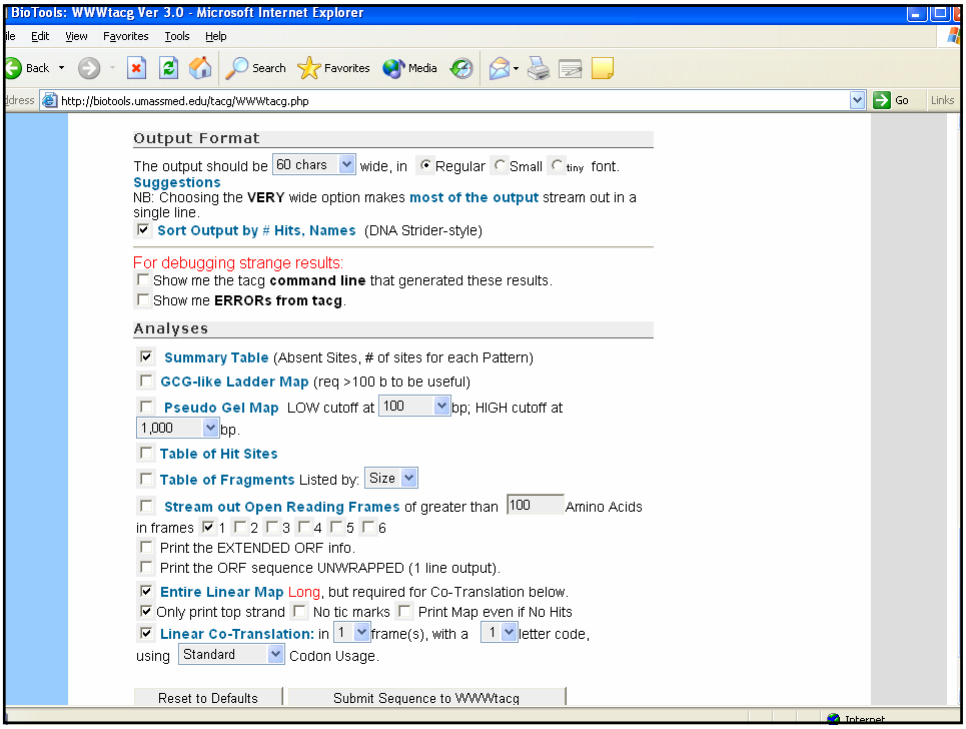
Biotools@UMASS Medical School

- http://biotools.umassmed.edu/tacg/WWWtacg.php
- Enter your DNA sequence
- Select limits



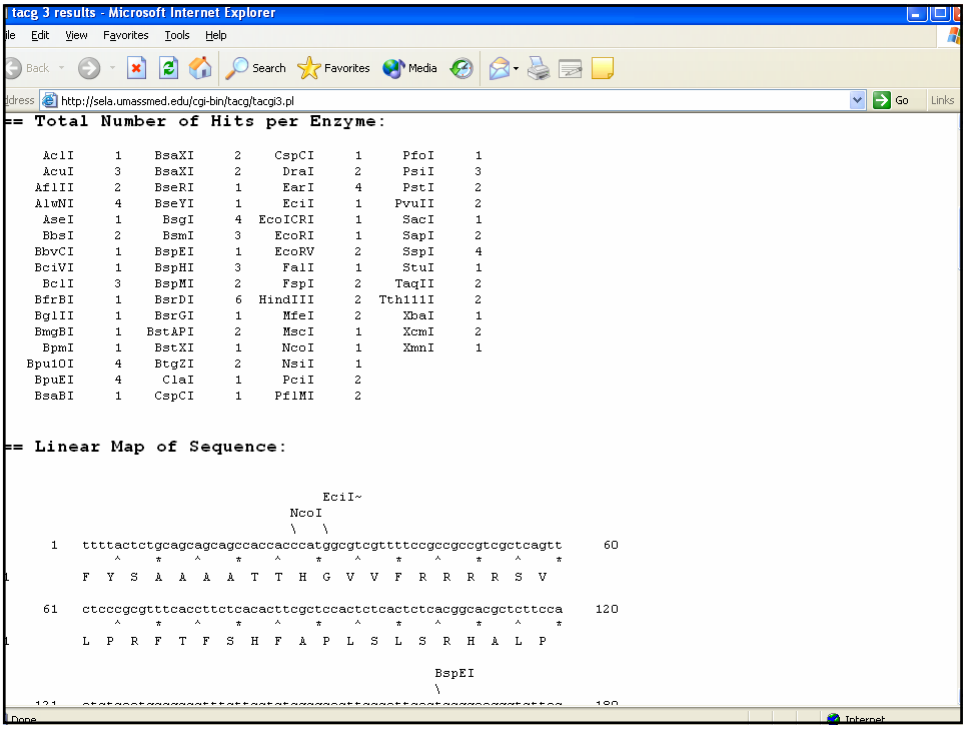

BioTools: WWWtacg Ver 3.0 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://biotools.umassmed.edu/tacg/WWWtacg.php
Go Links
--EITHER--Paste in the RAW sequence (>4 bp) in the window below...
Unlabeled
[Optional Output Label for pasted-in sequence]
2881 acgtgaattt tgatcgaaat gaatgatgtc gatggatatc cgaacacatg caaatttgta
2941 ccatctattt atttatcttg attttcttcg taggttattt gcttcttata acgttgtagt
3001 aatgtaagct tgtttctctt atcttttgtc tggaagaata ataagggaat ggaacagaac
3061 tccattaacc ttgattgtga gttacattgt aaaggaacgg aagtaaacag aatttattat
3121 aatttttaac ttcgctaact gtctttttaa taaaaaaaaa aaaaaa
--OR-- use the file browser below
NB: if there's sequence in the window above, the file will be ignored.
[Mac/Win32 versions of Netscape up to V4.7 (but not linux/Unix) will not accept more than ~30K chars in the paste-in buffer. If you need to load large files, save as text and load them via the file browser below]
Use the file browser below to upload a DNA sequence file (<500,000 characters) to send:
(Output will be labeled with the file name.)
Browse...
Force sequence file to be Read as 'raw'. (only numbers, spaces, and sequence)
Sequence Selection and Numbering Options
The sequence should be analyzed from bases 1 to END
(use "END" to indicate end; integers for other endpoints.)
The (sub)sequence should be analyzed as: Linear Circular
Sequence numbering should start from 1 (Only for Linear Map; Sites are numbered starting from 1.)


BioTools: WWWtacg Ver 3.0 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://biotools.umassmed.edu/tacg/WWWtacg.php
Go Links
Reset to Defaults
Submit Sequence to WWWtacg
Restriction Enzyme Selection
Select the Restriction Enzymes either By Characteristic or Explicitly
Filtering options below apply to entire REBASE files (yours or mine), but NOT to explicitly picked enzymes.
Search for possible SILENT Sites (supercedes all other selections and forces Linear Map Mode).
By Characteristic
This side is the default.
Overhang Don't Care
Number of cuts:
Minimum: 1
Maximum: 1
Magnitude of Recognition Sequence: 6 base pairs or greater.
Cost No Limit
Selections filtered from from the Latest REBASE file
--OR----------
from YOUR uploaded GCG-formatted REBASE file,
using the file browser below.
Uploads supercede the menu selection.
Explicitly
This side is selected if ANY names are selected
Simulate multiple digestion as well as singles.
AatII AccI AceIII AciI AclI AflII AflIII
BmgI BplI BpmI Bpu10I Bpu1102I BsaI BsaAI
BsrGI BssHII BssSI Bst4CI BstAPI BstDSI BstEII
EcoRV FauI Fnu4HI FokI FseI FspI GdiII
NdeI NgoAIV NheI NlaIII NlaIV
ScrFI SexAI SfaNI SfcI SfiI



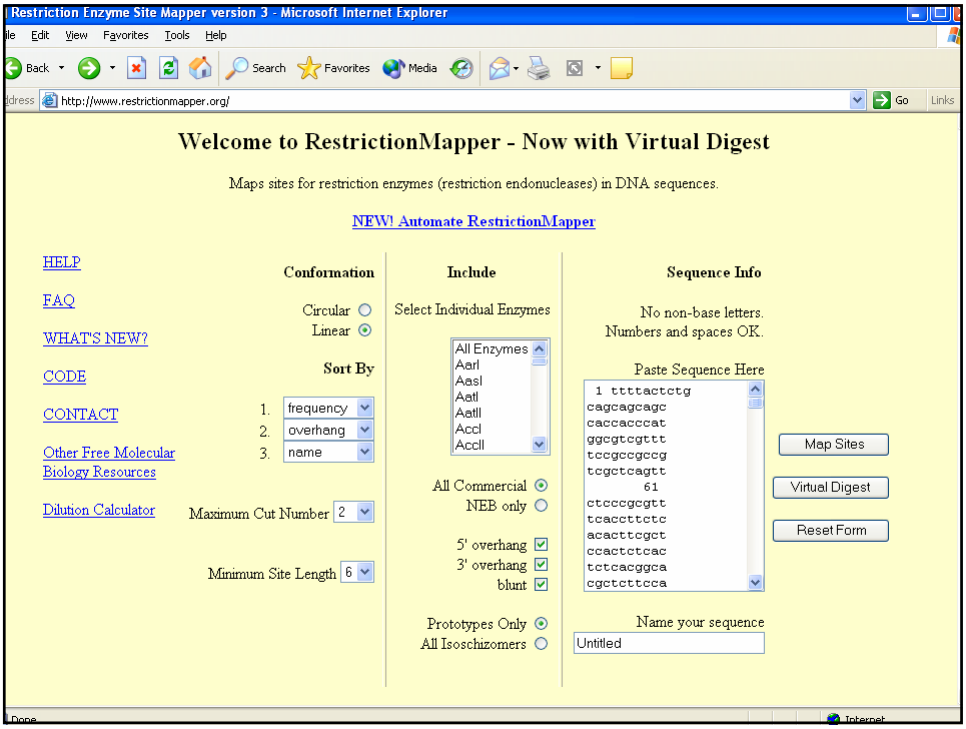
Results of query

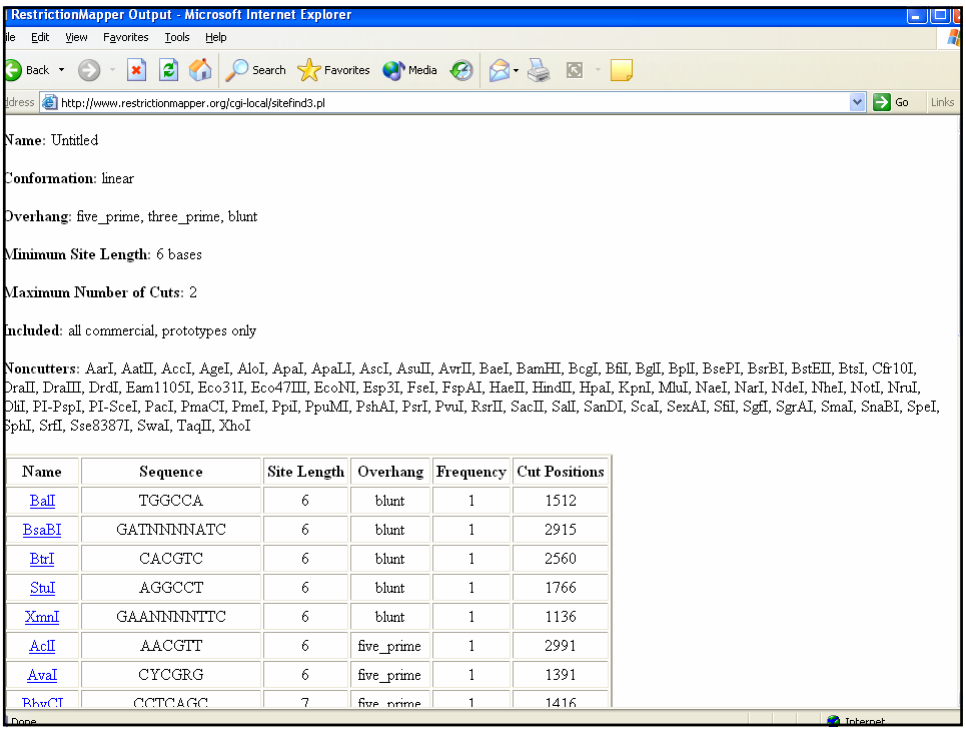
- Enzymes that DO NOT MAP
  - table
- Enzymes that cut and number of cuts
  - table
- Linear map of sequence with cut sites and amino acid translation



Restrictionmapper

- Locates restriction cut sites
- http://www.restrictionmapper.org/
- Does not draw map with DNA sequence


Restriction Enzyme Site Mapper version 3 - Microsoft Internet Explorer
http://www.restrictionmapper.org/
Welcome to RestrictionMapper - Now with Virtual Digest
Maps sites for restriction enzymes (restriction endonucleases) in DNA sequences.
NEW! Automate RestrictionMapper
HELP
FAQ
WHAT'S NEW?
CODE
CONTACT
Other Free Molecular Biology Resources
Dilution Calculator
Conformation
Circular
Linear
Sort By
1. frequency
2. overhang
3. name
Maximum Cut Number 2
Minimum Site Length 6
Include
Select Individual Enzymes
All Enzymes
AarI
AasI
AatI
AatII
AccI
AccII
All Commercial
NEB only
5' overhang
3' overhang
blunt
Prototypes Only
All Isoschizomers
Sequence Info
No non-base letters.
Numbers and spaces OK.
Paste Sequence Here
1 ttttactctg
cagcagcagc
caccacccat
ggcgtcgttt
tccgccgccg
tcgctcagtt
61
ctcccgcgtt
tcaccttctc
acacttcgct
ccactctcac
tctcacggca
cgctcttcca
Map Sites
Virtual Digest
Reset Form
Name your sequence
Untitled


RestrictionMapper Output - Microsoft Internet Explorer
http://www.restrictionmapper.org/cgi-local/sitefind3.pl
Name: Untitled
Conformation: linear
Overhang: five_prime, three_prime, blunt
Minimum Site Length: 6 bases
Maximum Number of Cuts: 2
Included: all commercial, prototypes only
Noncutters: AarI, AatII, AccI, AgeI, AloI, ApaI, ApaLI, AscI, AsuII, AvrII, BaeI, BamHI, BcgI, BfiI, BglI, BplI, BsePI, BsrBI, BstEII, BtsI, Cfr10I, DraII, DraIII, DrdI, Eam1105I, Eco31I, Eco47III, EcoNI, Esp3I, FseI, FspAI, HaeII, HindII, HpaI, KpnI, MluI, NaeI, NarI, NdeI, NheI, NotI, NruI, OliI, PI-PspI, PI-SceI, PacI, PmaCI, PmeI, PpiI, PpuMI, PshAI, PsrI, PvuI, RsrII, SacII, SalI, SanDI, ScaI, SexAI, SfiI, SgfI, SgrAI, SmaI, SnaBI, SpeI, SphI, SrfI, Sse8387I, SwaI, TaqII, XhoI
Name
Sequence
Site Length
Overhang
Frequency
Cut Positions
BalI
TGGCCA
6
blunt
1
1512
BsaBI
GATNNNNATC
6
blunt
1
2915
BtrI
CACGTC
6
blunt
1
2560
StuI
AGGCCT
6
blunt
1
1766
XmnI
GAANNNNTTC
6
blunt
1
1136
AclI
AACGTT
6
five_prime
1
2991
AvaI
CYCGRG
6
five_prime
1
1391
BbvCI
CCTCAGC
7
five_prime
1
1416

Web Mapping Tools

There are other some free mapping tools on the web for finding restriction sites and making text maps, but not for nice graphical maps

- WebÇutter (Max Heiman, Yale Univ.)
  http://www.firstmarket.com/cutter/cut2.html
- EMBOSS Restrict (EMBL Institut Pasteur)
  http://bioweb.pasteur.fr/seqanal/interfaces/restrict.html
- Restriction Maps (Colorado State Univ.)

http://arbl.cvmbs.colostate.edu/molkit/mapper/index.html
(uses Java)

WebCutter

Webcutter is a free on-line tool to restriction map nucleotide sequence (text output)

http://www.firstmarket.com/cutter/cut2.html

- Webcutter includes the option of finding restriction sites that can be introduced into a sequence by silent mutagenesis

```
NgoAIV                          Esp1396I
      MroNI                       AccB7I
aatggaagccggcggcacctcgctaacggattcaccactccaagaattggagccaatcaattcttgcggagaact   base pairs
ttaccttcggccgccgtggagcgattgcctaagtggtgaggttcttaacctcggttagttaagaacgcctcttga   1276 to 1350
      NgoMI                       PflMI
       NaeI                       Van91I

                 AviII     EcoT14I
      Mva1269I   StyI Esp1396I
      FspI       Eco130I                 GsuI
gtgaatgcgcaaaccaaccccttggcagaacatatccatcgcgtccgccatctccagcagccgcacgcggcgcatc   base pairs
cacttacgcgtttggttgggaaccgtcttgtataggtagcgcaggcggtagaggtcgtcggcgtgcgccgcgtag   1351 to 1425
      BsmI       ErhI PflMI                 BpmI
      BsaMI      BssT1I Van91I
      Acc16I     AccB7I
```

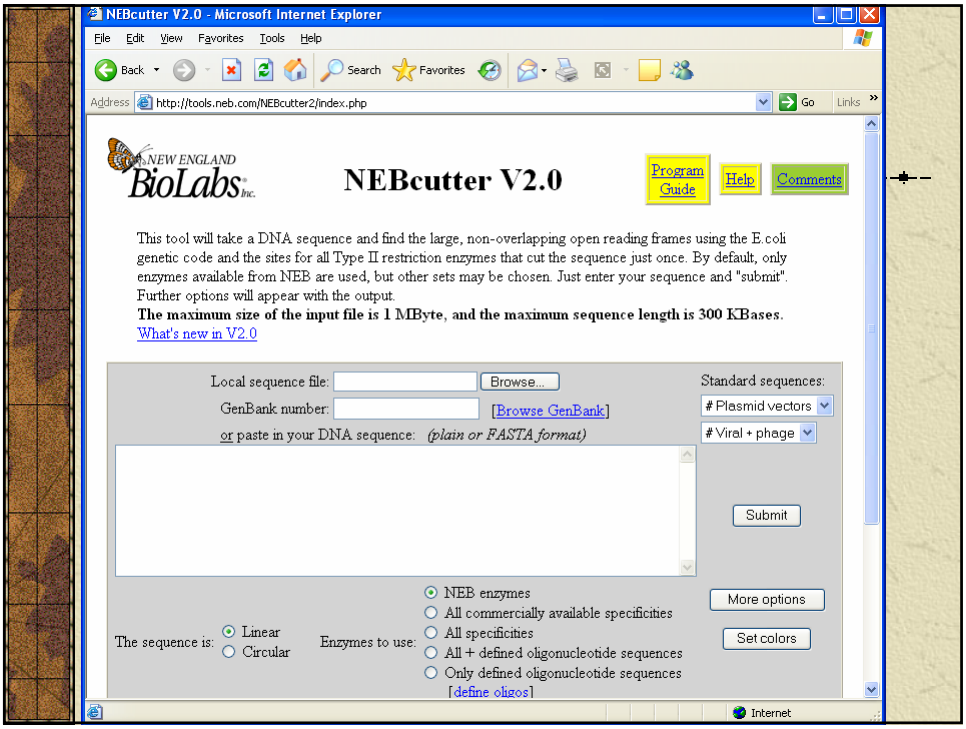
EMBOSS Restrict

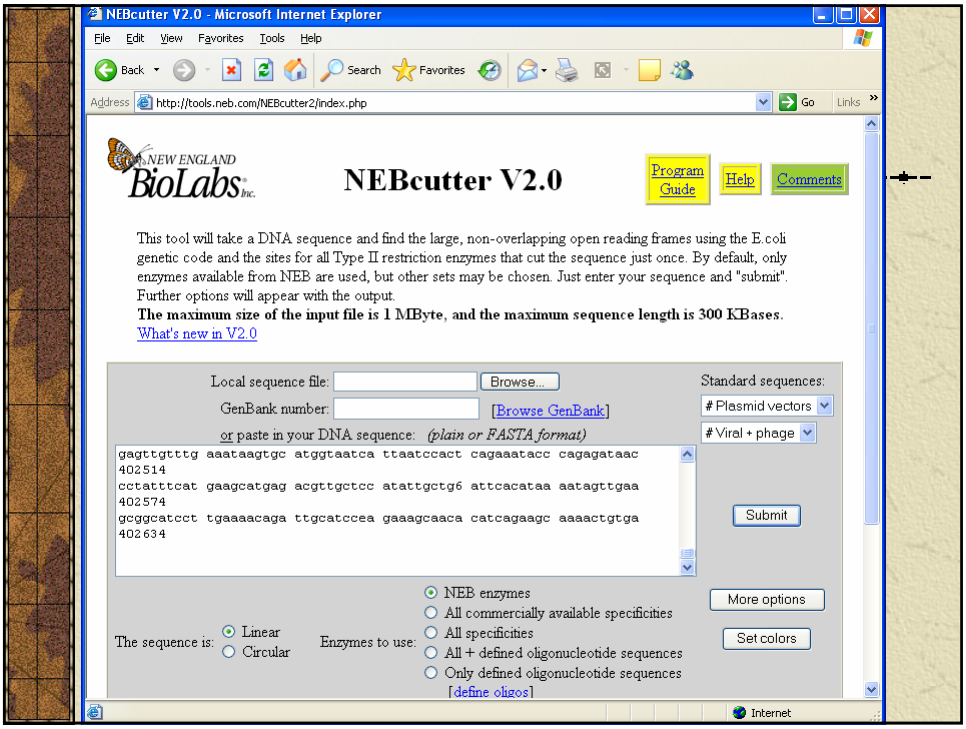
http://bioweb.pasteur.fr/seqanal/interfaces/restrict.html

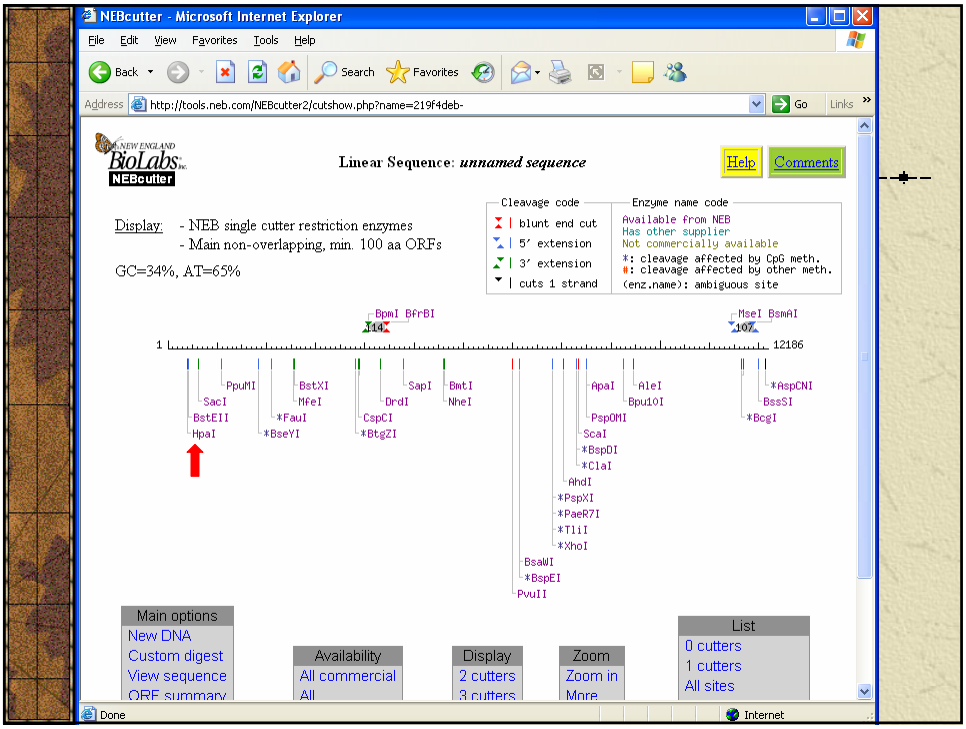
```
# Restrict of , from 1 to 150
# Minimum cuts per enzyme: 1
# Maximum cuts per enzyme: 2000000000
# Minimum length of recognition site: 4
# Blunt ends allowed     # Sticky ends allowed
# DNA is linear          # No ambiguities allowed
# Number of hits: 24
# Base Number   Enzyme          Site          5'     3'     [5'     3']
       8       NgoAIV          GCCGGC         8      12
       8       NaeI            GCCGGC         10     10
       9       MspI            CCGG           9      11
       15      AciI            CCGC           12     14
       18      MnlI            CCTC           28     27
       36      HphI            GGTGA          24     23
       45      Tsp509I         AATT           44     48
       59      Tsp509I         AATT           58     62
       69      AciI            CCGC           66     68
       78      BsmI            GAATGC         84     82
       81      FspI            TGCGCA         83     83
       82      Hin6I           GCGC           82     84
       82      HhaI            GCGC           84     82
       114     ThaI            CGCG           115    115
       119     AciI            CCGC           119    121
       119     HgaI            GACGC          104    109
       123     EciI            GGCGGA         108    106
       131     BbvI            GCAGC          143    147
       131     BpmI            CTGGAG         111    109
       135     AciI            CCGC           135    137
       140     ThaI            CGCG           141    141
       144     AciI            CCGC           141    143
       144     Hin6I           GCGC           144    146
       144     HhaI            GCGC           146    144
```

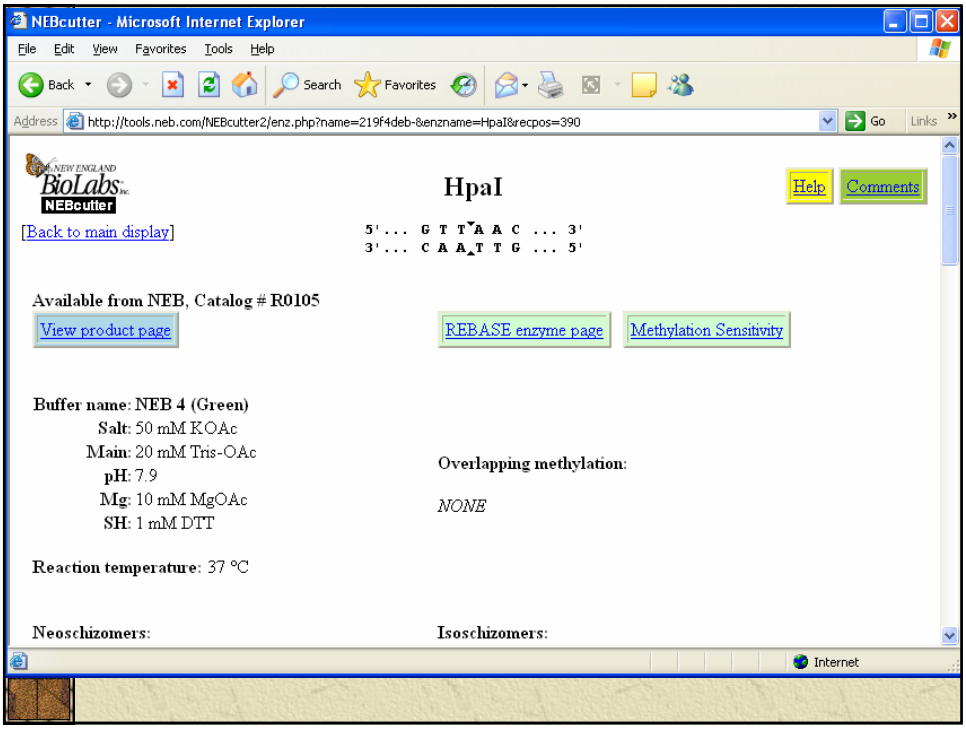
NEBcutter

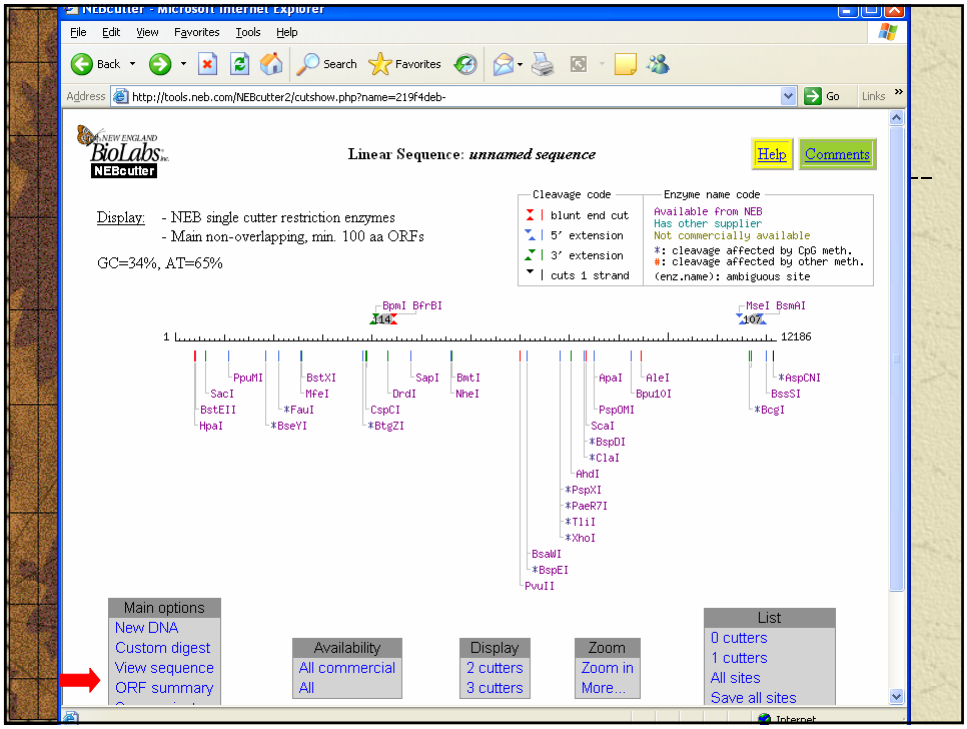
- New England BioLabs

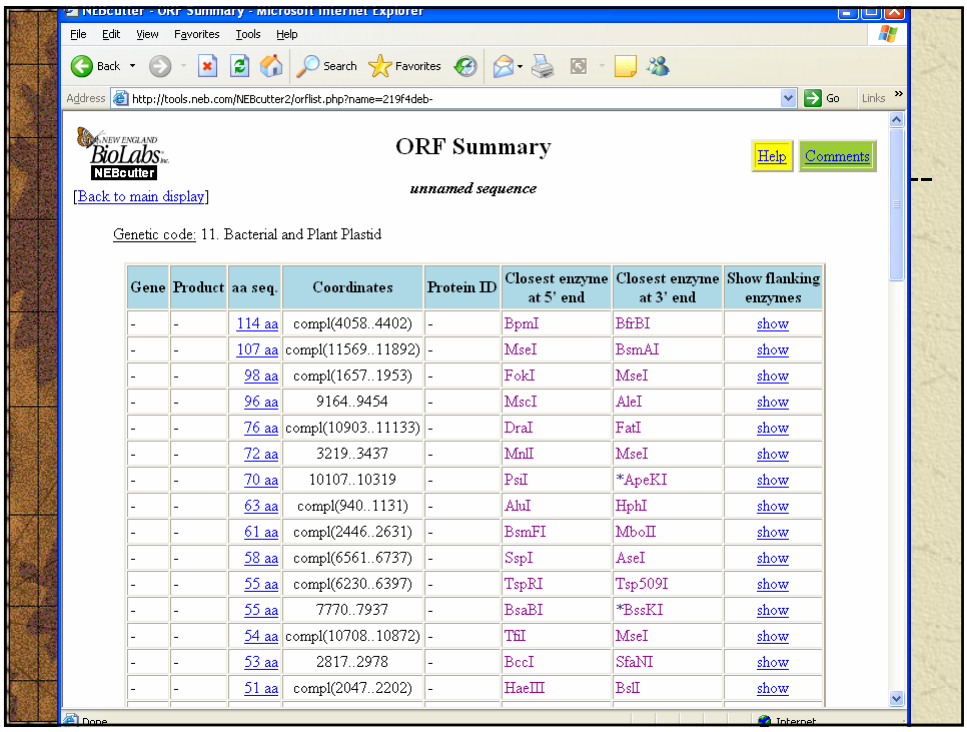

NEBcutter V2.0 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/index.php
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter V2.0
Program Guide
Help
Comments
This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output.
The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.
What's new in V2.0
Local sequence file:
Browse...
GenBank number:
[Browse GenBank]
or paste in your DNA sequence: (plain or FASTA format)
Standard sequences:
Plasmid vectors
Viral + phage
Submit
More options
Set colors
The sequence is:
Linear
Circular
Enzymes to use:
NEB enzymes
All commercially available specificities
All specificities
All + defined oligonucleotide sequences
Only defined oligonucleotide sequences
[define oligos]
Internet


NEBcutter V2.0 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/index.php
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter V2.0
Program Guide
Help
Comments
This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output.
The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.
What's new in V2.0
Local sequence file:
Browse...
GenBank number:
[Browse GenBank]
or paste in your DNA sequence: (plain or FASTA format)
gagttgtttg aaataagtgc atggtaatca ttaatccact cagaaatacc cagagataac
402514
cctatttcat gaagcatgag acgttgctcc atattgctg6 attcacataa aatagttgaa
402574
gcggcatcct tgaaaacaga ttgcatccea gaaagcaaca catcagaagc aaaactgtga
402634
Standard sequences:
Plasmid vectors
Viral + phage
Submit
More options
Set colors
The sequence is:
Linear
Circular
Enzymes to use:
NEB enzymes
All commercially available specificities
All specificities
All + defined oligonucleotide sequences
Only defined oligonucleotide sequences
[define oligos]
Internet

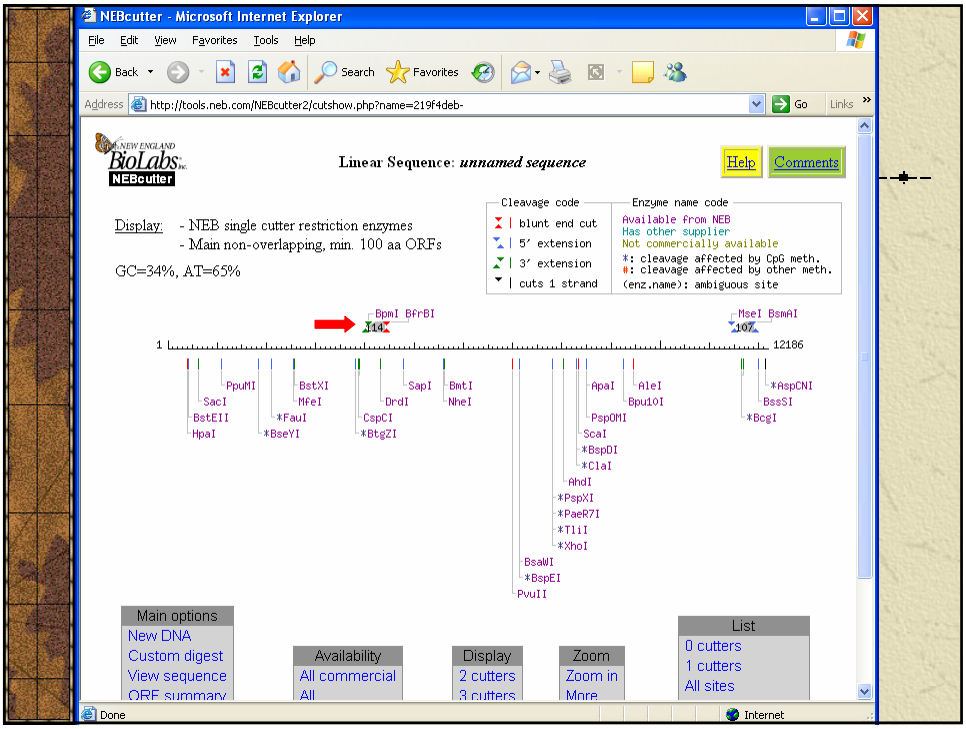

NEBcutter - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/cutshow.php?name=219f4deb-
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter
Linear Sequence: unnamed sequence
Help
Comments
Cleavage code
blunt end cut
5′ extension
3′ extension
cuts 1 strand
Enzyme name code
Available from NEB
Has other supplier
Not commercially available
*: cleavage affected by CpG meth.
#: cleavage affected by other meth.
(enz.name): ambiguous site
Display: - NEB single cutter restriction enzymes
- Main non-overlapping, min. 100 aa ORFs
GC=34%, AT=65%
BpmI BfrBI
MseI BsmAI
1
12186
PpuMI
SacI
BstEII
HpaI
BstXI
MfeI
*FauI
*BseYI
SapI
DrdI
CspCI
*BtgZI
BmtI
NheI
ApaI
PspOMI
ScaI
*BspDI
*ClaI
AhdI
*PspXI
*PaeR7I
*TliI
*XhoI
BsaWI
*BspEI
PvuII
AleI
Bpu10I
*AspCNI
BssSI
*BcgI
Main options
New DNA
Custom digest
View sequence
ORF summary
Availability
All commercial
All
Display
2 cutters
3 cutters
Zoom
Zoom in
More
List
0 cutters
1 cutters
All sites
Done
Internet


NEBcutter - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/enz.php?name=219f4deb-8enzname=HpaI&recpos=390
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter
HpaI
Help
Comments
[Back to main display]
5'... G T T A A C ... 3'
3'... C A A T T G ... 5'
Available from NEB, Catalog # R0105
View product page
REBASE enzyme page
Methylation Sensitivity
Buffer name: NEB 4 (Green)
Salt: 50 mM KOAc
Main: 20 mM Tris-OAc
pH: 7.9
Mg: 10 mM MgOAc
SH: 1 mM DTT
Overlapping methylation:
NONE
Reaction temperature: 37 °C
Neoschizomers:
Isoschizomers:
Internet

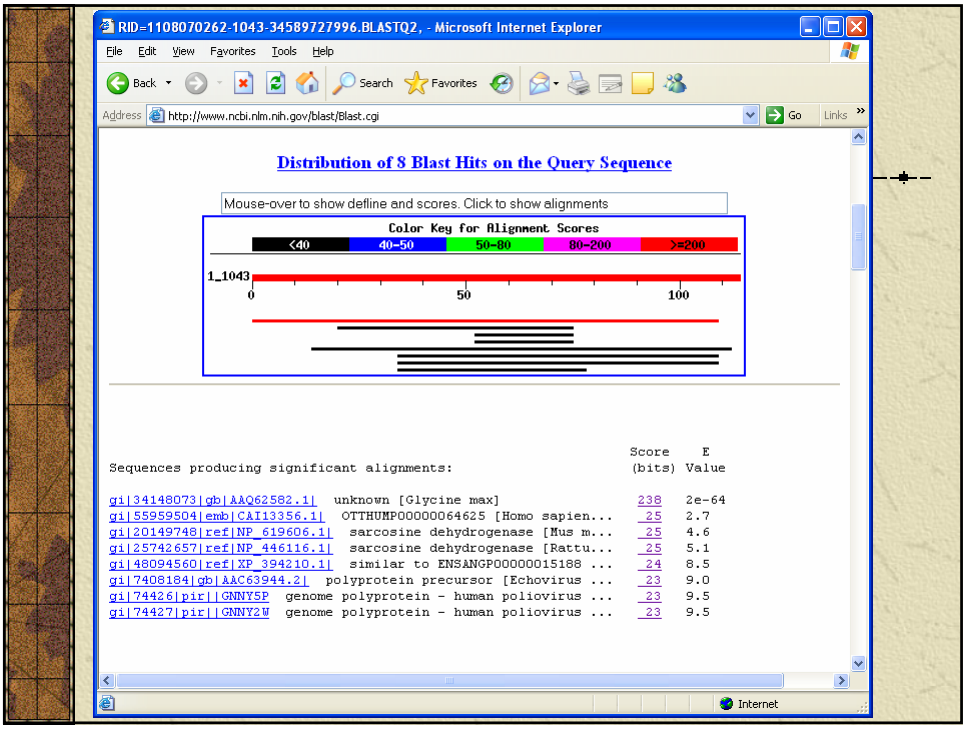


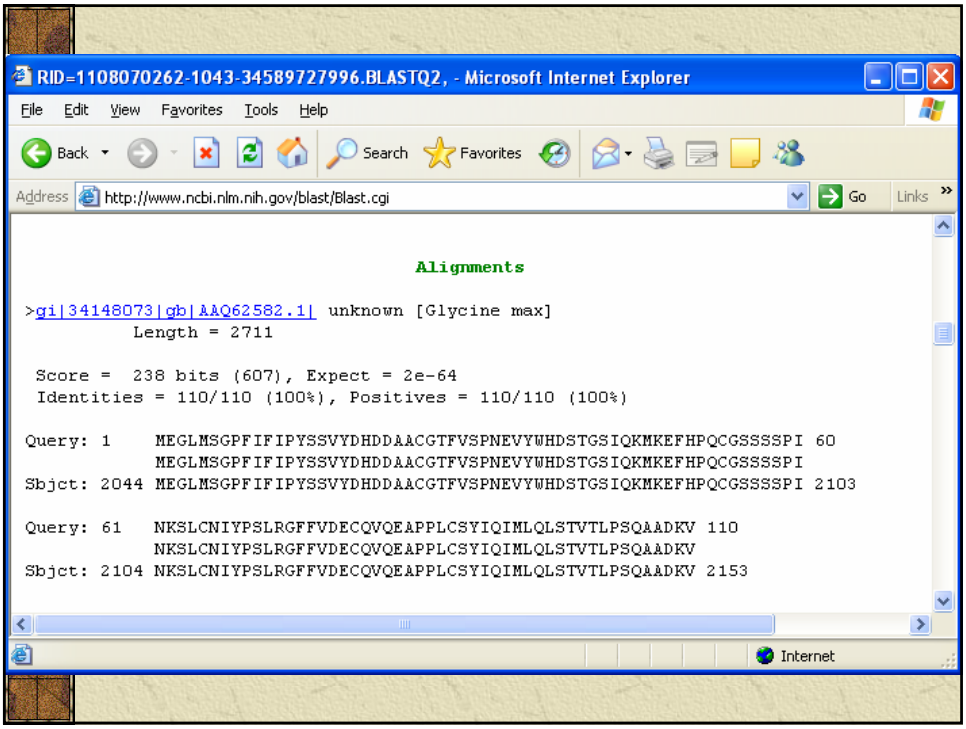


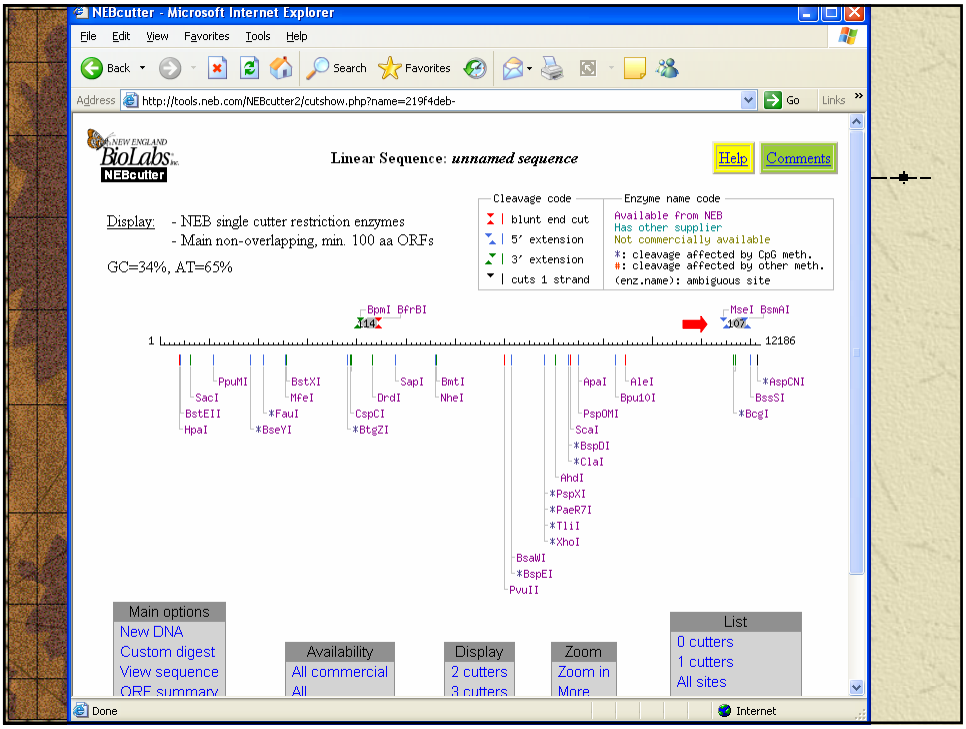
Gene	Product	aa seq.	Coordinates	Protein ID	Closest enzyme at 5' end	Closest enzyme at 3' end	Show flanking enzymes
-	-	114 aa	compl(4058..4402)	-	BpmI	BfrBI	show
-	-	107 aa	compl(11569..11892)	-	MseI	BsmAI	show
-	-	98 aa	compl(1657..1953)	-	FokI	MseI	show
-	-	96 aa	9164..9454	-	MscI	AleI	show
-	-	76 aa	compl(10903..11133)	-	DraI	FatI	show
-	-	72 aa	3219..3437	-	MnlI	MseI	show
-	-	70 aa	10107..10319	-	PsiI	*ApeKI	show
-	-	63 aa	compl(940..1131)	-	AluI	HphI	show
-	-	61 aa	compl(2446..2631)	-	BsmFI	MboII	show
-	-	58 aa	compl(6561..6737)	-	SspI	AseI	show
-	-	55 aa	compl(6230..6397)	-	TspRI	Tsp509I	show
-	-	55 aa	7770..7937	-	BsaBI	*BssKI	show
-	-	54 aa	compl(10708..10872)	-	TfiI	MseI	show
-	-	53 aa	2817..2978	-	BccI	SfaNI	show
-	-	51 aa	compl(2047..2202)	-	HaeIII	BslI	show

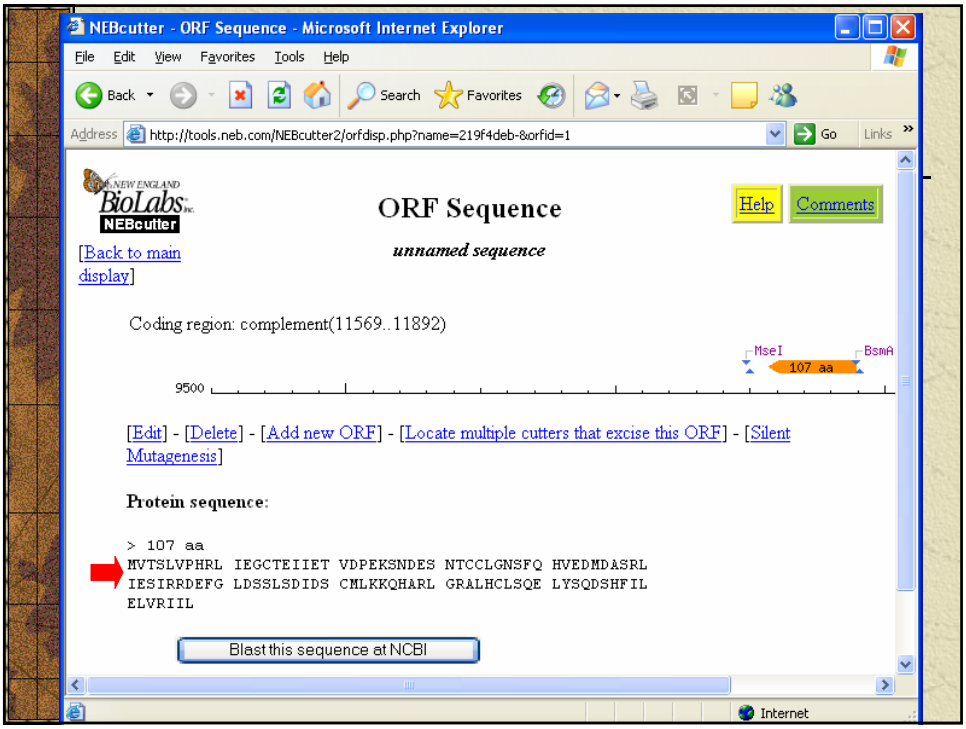

NEBcutter - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/cutshow.php?name=219f4deb-
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter
Linear Sequence: unnamed sequence
Help
Comments
Display: - NEB single cutter restriction enzymes
- Main non-overlapping, min. 100 aa ORFs
GC=34%, AT=65%
Cleavage code
blunt end cut
5′ extension
3′ extension
cuts 1 strand
Enzyme name code
Available from NEB
Has other supplier
Not commercially available
*: cleavage affected by CpG meth.
#: cleavage affected by other meth.
(enz.name): ambiguous site
BpmI BfrBI
114
MseI BsmAI
107
1
12186
PpuMI
SacI
BstEII
HpaI
BstXI
MfeI
*FauI
*BseYI
SapI
DrdI
CspCI
*BtgZI
BmtI
NheI
ApaI
PspOMI
ScaI
*BspDI
*ClaI
AhdI
*PspXI
*PaeR7I
*TliI
*XhoI
BsaWI
*BspEI
PvuII
AleI
Bpu10I
*AspCNI
BssSI
*BcgI
Main options
New DNA
Custom digest
View sequence
ORF summary
Availability
All commercial
All
Display
2 cutters
3 cutters
Zoom
Zoom in
More
List
0 cutters
1 cutters
All sites
Done
Internet

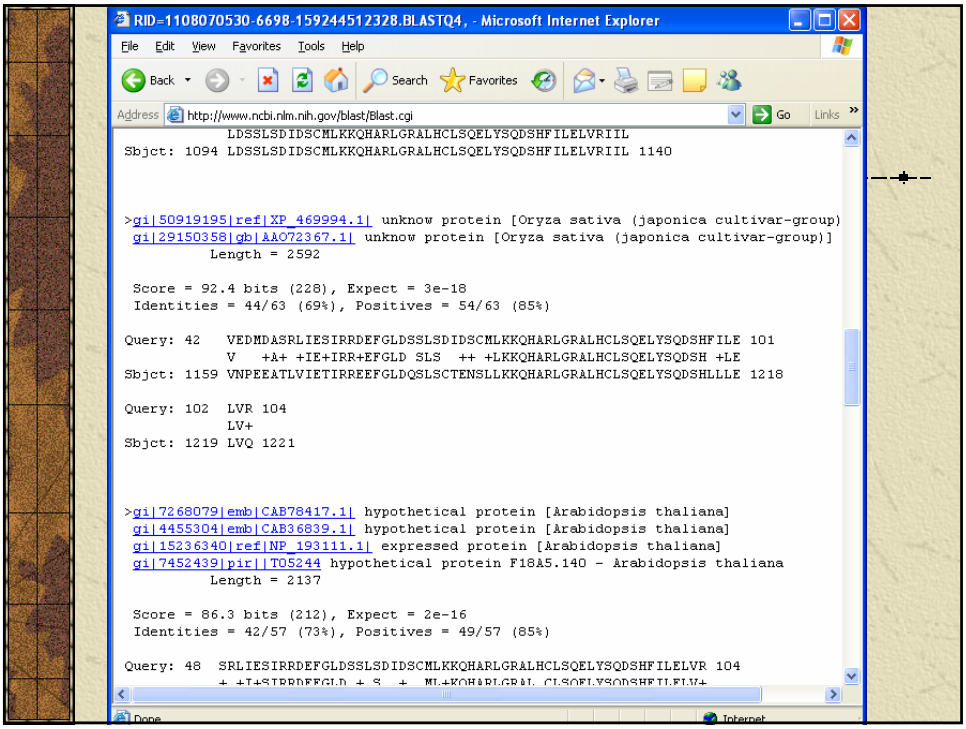

NEBcutter - ORF Sequence - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/orfdisp.php?name=219f4deb-&orfid=0
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter
ORF Sequence
Help
Comments
unnamed sequence
[Back to main display]
Coding region: complement(4058..4402)
BpmI BfrBI
114 aa
2000
6500
[Edit] - [Delete] - [Add new ORF] - [Locate multiple cutters that excise this ORF] - [Silent Mutagenesis]
Protein sequence:
> 114 aa
MEGLMSGPFI FIPYSSVYDH DDAACGTFVS PNEVYWHDST GSIQKMKEFH
PQCGSSSSPI NKSLCNIYPS LRGFFVDECQ VQEAPPLCSY IQIMLQLSTV
TLPSQAADKV SSEL
Blast this sequence at NCBI
Done
Internet


RID=1108070262-1043-34589727996.BLASTQ2, - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
Go Links
Distribution of 8 Blast Hits on the Query Sequence
Mouse-over to show defline and scores. Click to show alignments
Color Key for Alignment Scores
<40 40-50 50-80 80-200 >=200
1_1043
0 50 100
Sequences producing significant alignments: Score (bits) E Value
gi|34148073|gb|AAQ62582.1| unknown [Glycine max] 238 2e-64
gi|55959504|emb|CAI13356.1| OTTHUMP00000064625 [Homo sapien... 25 2.7
gi|20149748|ref|NP_619606.1| sarcosine dehydrogenase [Mus m... 25 4.6
gi|25742657|ref|NP_446116.1| sarcosine dehydrogenase [Rattu... 25 5.1
gi|48094560|ref|XP_394210.1| similar to ENSANGP00000015188 ... 24 8.5
gi|7408184|gb|AAC63944.2| polyprotein precursor [Echovirus ... 23 9.0
gi|74426|pir||GNNY5P genome polyprotein - human poliovirus ... 23 9.5
gi|74427|pir||GNNY2W genome polyprotein - human poliovirus ... 23 9.5
Internet


RID=1108070262-1043-34589727996.BLASTQ2, - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
Go Links
Alignments
>gi|34148073|gb|AAQ62582.1| unknown [Glycine max]
Length = 2711
Score = 238 bits (607), Expect = 2e-64
Identities = 110/110 (100%), Positives = 110/110 (100%)
Query: 1 MEGLMSGPFIFIPYSSVYDHDDAACGTFVSPNEVYWHDSTGSIQKMKEFHPQCGSSSSPI 60
MEGLMSGPFIFIPYSSVYDHDDAACGTFVSPNEVYWHDSTGSIQKMKEFHPQCGSSSSPI
Sbjct: 2044 MEGLMSGPFIFIPYSSVYDHDDAACGTFVSPNEVYWHDSTGSIQKMKEFHPQCGSSSSPI 2103
Query: 61 NKSLCNIYPSLRGFFVDECQVQEAPPLCSYIQIMLQLSTVTLPSQAADKV 110
NKSLCNIYPSLRGFFVDECQVQEAPPLCSYIQIMLQLSTVTLPSQAADKV
Sbjct: 2104 NKSLCNIYPSLRGFFVDECQVQEAPPLCSYIQIMLQLSTVTLPSQAADKV 2153
Internet


NEBcutter - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/cutshow.php?name=219f4deb-
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter
Linear Sequence: unnamed sequence
Help Comments
Display: - NEB single cutter restriction enzymes
- Main non-overlapping, min. 100 aa ORFs
GC=34%, AT=65%
Cleavage code
blunt end cut
5′ extension
3′ extension
cuts 1 strand
Enzyme name code
Available from NEB
Has other supplier
Not commercially available
*: cleavage affected by CpG meth.
#: cleavage affected by other meth.
(enz,name): ambiguous site
BpmI BfrBI
MseI BsmAI
1
12186
PpuMI
SacI
BstEII
HpaI
BstXI
MfeI
*FauI
*BseYI
SapI
DrdI
CspCI
*BtgZI
BmtI
NheI
ApaI
AleI
Bpu10I
PspOMI
ScaI
*BspDI
*ClaI
AhdI
*PspXI
*PaeR7I
*TliI
*XhoI
BsaWI
*BspEI
PvuII
*AspCNI
BssSI
*BcgI
Main options
New DNA
Custom digest
View sequence
ORF summary
Availability
All commercial
All
Display
2 cutters
3 cutters
Zoom
Zoom in
More
List
0 cutters
1 cutters
All sites
Done
Internet


NEBcutter - ORF Sequence - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://tools.neb.com/NEBcutter2/orfdisp.php?name=219f4deb-&orfid=1
Go Links
NEW ENGLAND BioLabs Inc.
NEBcutter
ORF Sequence
Help Comments
[Back to main display]
unnamed sequence
Coding region: complement(11569..11892)
MseI
BsmA
107 aa
9500
[Edit] - [Delete] - [Add new ORF] - [Locate multiple cutters that excise this ORF] - [Silent Mutagenesis]
Protein sequence:
> 107 aa
MVTSLVPHRL IEGCTEIIET VDPEKSNDES NTCCLGNSFQ HVEDMDASRL
IESIRRDEFG LDSSLSDIDS CMLKKQHARL GRALHCLSQE LYSQDSHFIL
ELVRIIL
Blast this sequence at NCBI
Internet



Primer Design

- Polymerase chain reaction (PCR)
- DNA sequencing
- Cloning

PCR Primer Design

- The design of PCR (and sequencing) primers is relatively simple from a computational point of view: just search along a sequence and find short sub-sequences that fit certain criteria.
- However, since the molecular biology of PCR is very complex, the nature of these criteria is not at all obvious.
- All primers design software uses approx-imately the same criteria and computing algorithms. Graphical output is not necessary.

Molecular Biology of PCR

- The fundamental Molecular Biology of PCR is not well understood.
- We know what happens in a descriptive sense, but not the physical chemistry/thermodynamics
- The rules for choosing PCR primers are a rough combination of educated guesses and old fashioned trial-and-error.
- None of the published formulas for calculating annealing temperatures has been proven to give better than a rough estimate-however, most work!

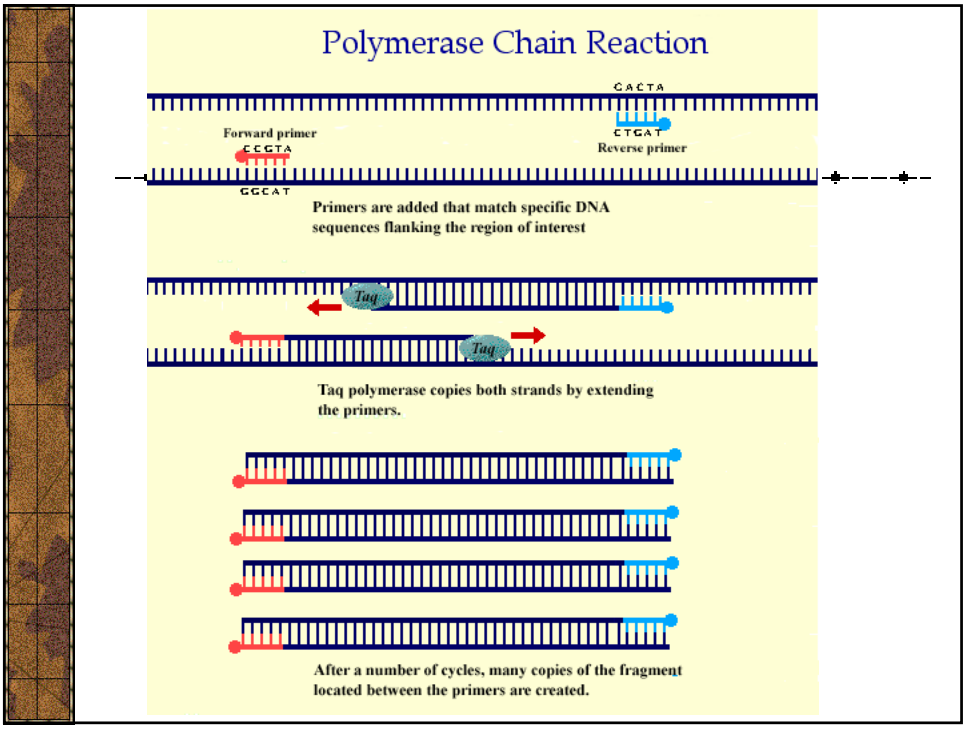
Polymerase chain reaction

- Amplify DNA fragment for cloning
- RT-PCR to measure gene expression
- Detect polymorphisms for mapping or diagnostics

- Need to design primers from DNA sequence

How PCR Works

- DNA and two primers are combined in a salt solution with dNTPs and a heat stable DNA polymerase enzyme
- The primers match some sequence in the target DNA
- The solution is rapidly heated to DNA denaturing temperatures (~95°C) and cooled to a temperature where the polymerase can function
- Each thermal cycle generates copies of the sequence between the primers, so the total number of fragments amplifies in an exponential fashion: 2, 4, 8,16, 32, 64, etc.

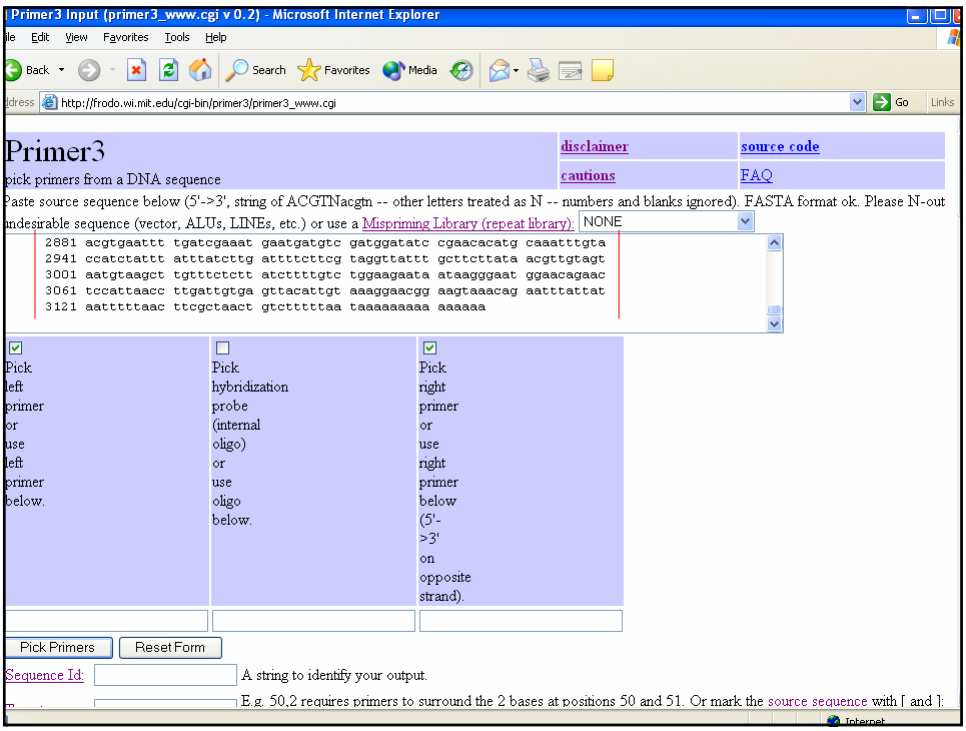


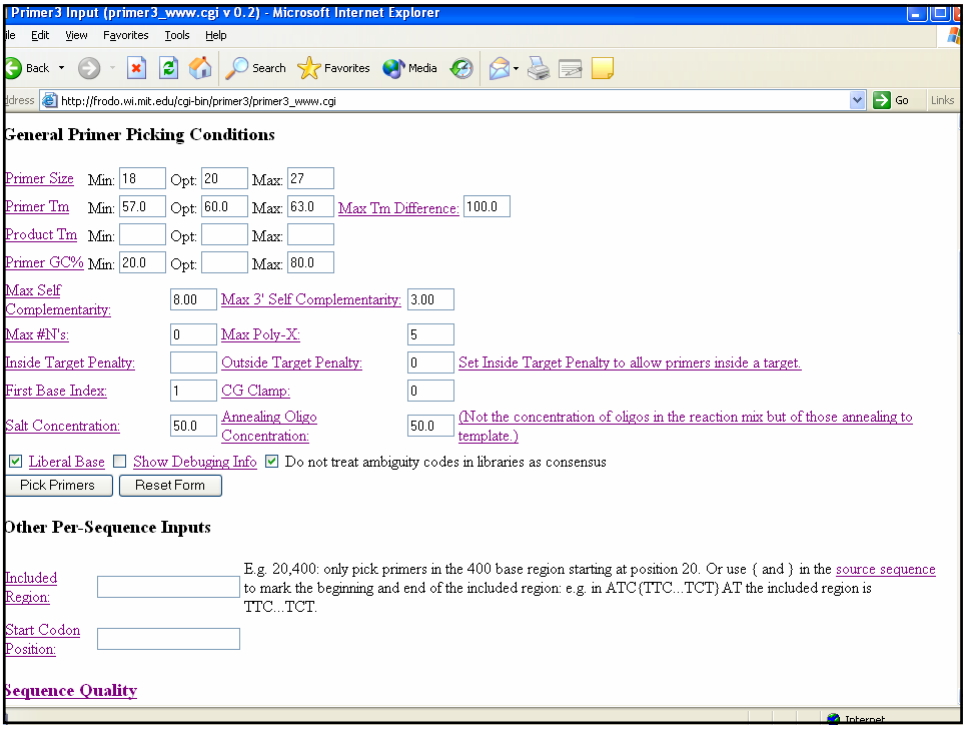
Primer Design Rules

- primers should be at least 15 base pairs long
- have at least 50% G/C content
- anneal at a temperature in the range of 50-65 degrees C
- Usually higher annealing temperatures (Tm) are better (i.e. more specific for your desired target)
- forward and reverse primer should anneal at approximately the same temperature

Primer design sites

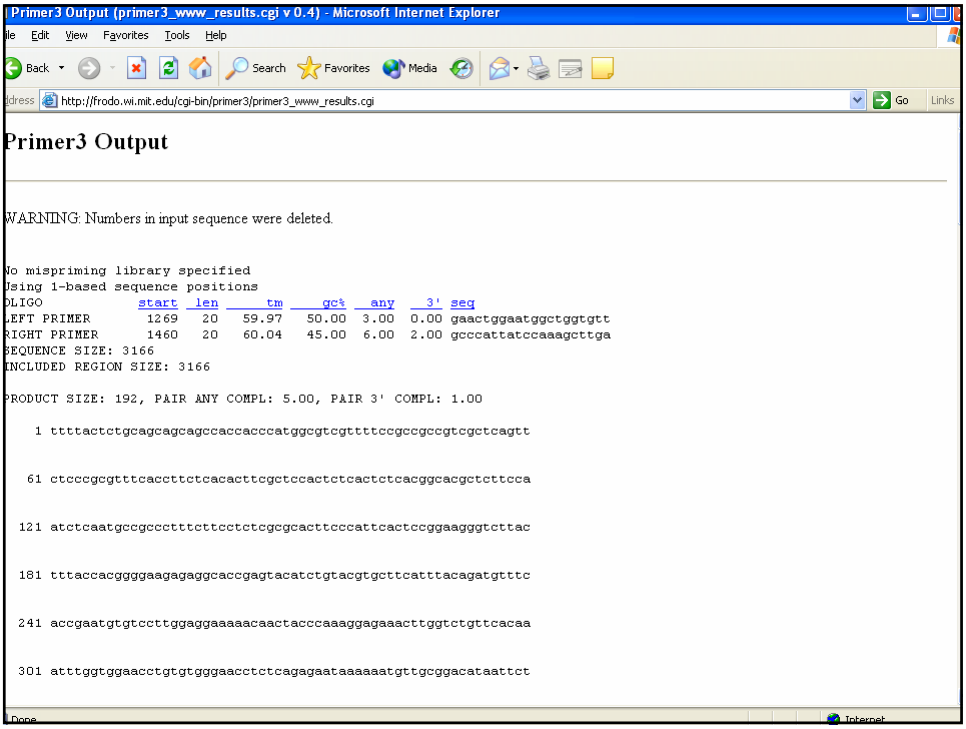
- Primer design sites
  - http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
- Primer3
  - http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
- GenomeWeb
  - http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-primer.html
  - Web Primer





Results

- Provides best and four other primer pairs
- Primer sequences
- Start nt for each primer
- Length of each primer
- Tm
- GC content
- Size of PCR product



Primer Problems

- primers should flank the sequence of interest
- primer sequences should be unique
- primers that match multiple sequences will give multiple products
- repeated sequences can be amplified - but only if unique flanking regions can be found where primers can bind
- primers can have self-annealing regions within each primer (i.e. hairpin and foldback loops)
- pairs of primers can anneal to each other to form "primer dimers"

Differential Primers

- New challenges for PCR primer design
  - gene-specific primers (for multi-gene families)
  - identify specific species or strains of organisms
    - molecular diagnostics/detectors
  - Consensus primers
    - amplify a gene from all of a diverse group of organsims (eg. bacterial 16-S rDNA)
- Need to work with multiple alignments and find differential or conserved regions

Primer Design on the Web

- There are a bunch of good PCR primer design programs on the web:
  - **Primer 3** at the MIT Whitehead Institute

http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

  - **Cassandra** at the Univ. of Southern California

http://www-hto.usc.edu/software/procrustes/cassandra/cass_frm.html

  - **GeneFisher** by Folker Meyer & Chris Schleiermacher at Bielefeld University, Germany

http://bibiserv.TechFak.Uni-Bielefeld.DE/genefisher/

  - **Xprimer** at the Virtual Genome Center, Univ. Minnesota Medical School

http://alces.med.umn.edu/rawprimer.html

Other Technologies

- Multiplex PCR
  - GeneScan (ABI)
- PCR related technologies
  - Primer extension
    - Taqman (ABI)
    - Orchid
    - Pyrosequencing
  - Ligase chain reaction
  - Oligos for microarrays
  - DNA sequencing reactions

RNA secondary structure

- Look for secondary structures

What we covered today

- Restriction site analysis
  - Cloning
  - Southern blot analysis
- Primer design
  - Cloning, amplification gene expression, sequencing